

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
15 August 2002 (15.08.2002)

PCT

(10) International Publication Number
WO 02/063420 A2

(51) International Patent Classification⁷: G06F
(21) International Application Number: PCT/US02/01763
(22) International Filing Date: 23 January 2002 (23.01.2002)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
09/777,807 5 February 2001 (05.02.2001) US

(71) Applicant: PREDICTIVE NETWORKS, INC.
[US/US]; Suite 200, 689 Massachusetts Avenue, Cambridge, MA 02139 (US).

(72) Inventors: HOSEA, Devin, F.; 3 Gloucester Street, Boston, MA 02115 (US). ZIMMERMAN, Richard, Scott; 22 Cross Street, Belmont, MA 02478 (US). RASCON, Arthur, P.; 425 Woburn Street, Apt #47, Lexington, MA 02420 (US). ODDO, Anthony, Scott; 90 Wenham Street #3, Jamaica Plain, MA 02130 (US). THURSTON,

Nathaniel, John; 68 Pearson Road, #2, Somerville, Massachusetts 02144 (US).

(74) Agents: LARI, Ayla, A. et al.; Hale and Dorr LLP, 60 State Street, Boston, MA 02109 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

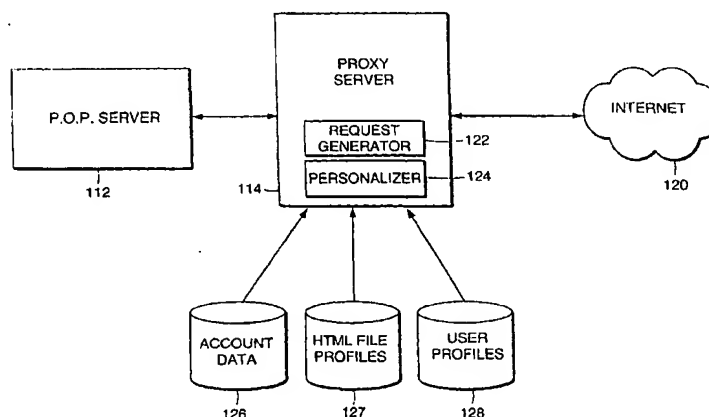
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR WEB PAGE PERSONALIZATION



(57) Abstract: The invention includes a method and system for personalizing displays of published Web pages provided by Web content providers to meet the interests of Web users accessing the pages, based on profiles of the users. The system preferably provides to the requesting user, through a proxy server, an edited version of the HTML file for the original published Web page that is served by a host Web server. The system uses profiles that may include demographic and psychographic data to edit the requested Web page. The content of a Web page as published by a host Web server may be coded to correlate components of the Web page with demographic and psychographic data. The user profiles may then be used to filter the content of a coded Web page for delivery to a requesting user. The system may rearrange content on a published Web page so that content determined to be of higher interests to a user is more prominently featured or more easily or quickly accessible. The system may also delete content on a published Web page that is determined to be of low interest to a user. In embodiments of the invention, a single proxy server or proxy server system personalizes Web pages from multiple Web servers, using a single user profile for a user.

WO 02/063420 A2

WO 02/063420 A2



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Method And System For Web Page Personalization

Field of the Invention

5

The invention relates generally to systems and methods for targeting World Wide Web ("Web") content to interested users and specifically to systems and methods for automatically personalizing delivered Web pages based on the preferences of the users requesting the Web pages.

10

Background of the Invention

15

20

25

Many Web sites attempt to catalog or provide access to an enormous amount of material, typically presented through Web "pages," in a multiplicity of subject areas or categories. For example, an Internet "portal" or "search engine" Web site, designed to help users find the Web content that is of interest to them, may list or otherwise incorporate millions of Web sites and/or individual Web pages pertaining to thousands of subject areas, such as Arts, Computers, Sports, Entertainment, etc. Also by way of example, a retail Web site or "e-tailer" may offer products in a number of categories, such as Women's Clothing, Men's Clothing, Household Appliances, Lawn and Garden Products, etc. In order to provide access to such large amounts of diverse material, such Web sites typically initially present information on a home page or other high-level pages that lead to a variety of content and subject areas. Because these pages are entry points to a Web site for a diverse, anonymous group of users, these high-level Web pages are typically designed for universal appeal and convenience, with a generic organization. This approach allows users to then select and navigate to Web pages that cover the subject areas or categories of interest to them. These high-level pages typically have a generic design for the further reason that once a Web page is published, i.e., available for Web users to access, the page,

including the content and the display format, is typically static (except in many cases for the advertising that displays with the page). Consequently, these pages may be designed to appeal to and suit the needs of the widest, most general group of users possible.

5 For a particular user to locate material of interest through a high-level Web page, he or she may have to scroll or scan through long lists of links to available material or link through several successive levels of increasing specificity. Users may find sifting through the large amount of available material using these methods to be slow, inefficient and cumbersome. Users may devote a considerable amount of time
10 simply to locating material of interest, and may miss such material altogether due to the prominence or predominance of other material. For example, if material of interest is "below the fold," i.e., requires scrolling after the Web page arrives in order to be visible, or if it is buried in a large amount of irrelevant information, a user may never consider it.

15 Moreover, the amount of information and content available on the Internet continues to grow at a fast pace. Not only are new Web sites being created every day, but existing Web sites continue to add new pages with new content. Web sites are reformatted and reorganized, so that users cannot rely on finding the same information in the same place twice consistently. The proliferation of Web content
20 makes it increasingly difficult for users to find what they are looking for. For example, Web portals may become less effective as the amount of Web content classified in their taxonomies increases. The same is true for virtually any large Web site.

If a user becomes frustrated with his or her inability to find desired material at
25 a particular Web site, the user is more likely to go to another Web site for that

material, and also to prefer that other Web site for future needs. As many users have the same experience with that Web site, they will similarly favor other Web sites. Eventually, the cumulative effect of these defections will be a significant reduction in traffic at the disfavored Web site. This trend will, in turn, reduce the ability of the disfavored Web site to generate sales and/or advertising revenue (the primary means of revenue generation for many non-e-tailing Web sites).

A more effective means of presenting Web content is tailoring Web content delivered to an individual user to meet the needs, preferences and interests of that user. Personalizing Web content delivered to users may generally improve user satisfaction. Typically on a Web-site-by-Web-site basis, some Web sites support personalization of some features of that particular Web site or a portion of that Web site. A user's experience with Web content on a Web page may include at least three components: content, layout, and graphics components. A Web site may allow a user to explicitly specify certain personalization options with respect to these components. A user may, for example, select desired types of content, perhaps by filling out a questionnaire or checklist.

For example, Yahoo!® (home page: www.yahoo.com), a well-known Web portal, includes a section called "My Yahoo!®" that allows a user to personalize some aspects of his or her interface to Yahoo!®. Figure 1 shows a screen-shot of the Yahoo!® home page 10. Figure 2 shows a screen-shot of the My Yahoo!® home page 30. My Yahoo!® allows a user to develop a "Front Page" directed to his or her interests. Figure 3 is a screen-shot of a Web page form 32 that allows a user to personalize the content of his or her Front Page by explicitly selecting desired content modules from a checklist 34. My Yahoo!® also allows a user to tailor the layout and the presentation features such as color and background based on his or her

preferences. Figure 4 is a screen-shot of a Web page 40 that allows a user to select a layout of the selected content modules for the Front Page; Figure 5 is a screen-shot of a Web page 50 that allows a user to select a particular color scheme for My Yahoo!® pages. Figure 6 is a screen-shot of a sample Front Page Web page 60 that has been personalized by a sports fan.

This personalization scheme has limited effectiveness. In the My Yahoo!® type of personalization scheme, the user explicitly specifies his or her preferences and, once specified, these recorded preferences typically remain the same unless explicitly updated or changed. These preferences are invoked by entering a user name and password for or at the particular Web site. Moreover, these preferences are generally specific to a local environment; for example, these preferences may be limited to My Yahoo!® and may not carry over to the “public areas” of the Yahoo!® Web site, let alone to other Web sites. Moreover, these preferences are not applicable to published Web content. In order to specify similar preferences on a different or unrelated Web site, the user must re-specify these preferences, if a personalization option is offered at all. Also, these preferences may not necessarily reflect how a user actually uses the Web. A user may select a content module related to Entertainment, but may not otherwise use the Web to access entertainment-related Web sites or purchase tickets. Thus, this type of personalization may not be useful for generalizing to other contexts.

U.S. Patent No. 6,128,655 to Fields, et al. shows the use of a proxy server that recasts published Web content from multiple Web sites in the look and feel of a hosting site for delivery to a requesting client. Although a user may choose a look and feel format by registering his or her preferences, the Web content is not personalized.

A need exists for a method and system for tailoring published Web page content in real-time, based on the user profiles of the users requesting the Web pages.

A need also exists for a method and system for personalizing published Web page content based on user profiles that accurately reflect Web use. A need also exists for a method and system for personalizing published Web page content from a number of Web sites using a single user profile for each user. A need also exists for a method and system that uses a proxy server system for personalizing published Web page content from a number of Web sites based on the user profiles of the users requesting the Web pages.

The present application is related to Utility Application Ser. No. 09/558,755 (“the ‘755 application”), entitled “Method and System for Web User Profiling and Selective Content Delivery,” filed April 21, 2000, which has a common assignee with the present application, and which is incorporated herein by this reference. The ‘755 application discloses, *inter alia*, a method and system for developing profiles for Web users that may be used in conjunction with the present invention.

Summary of the Invention

The present invention is directed to providing personalization of Web content in real-time to meet the interests of individual Web users. The invention includes a method and system for personalizing displays of published Web pages provided by Web content providers to meet the interests of Web users accessing the pages, based on profiles of the users. When a published Web page is requested by a user, the system arranges the constituent components of the requested Web page to better suit the interests of that user. In one aspect of the invention, the system rearranges content on a published Web page so that content determined to be of higher interest to a user is more prominently featured or more easily or quickly accessible. In another aspect

of the invention, the system edits content on a published Web page so that content determined to be of low interest to a user is eliminated.

The system uses user profiles that may include demographic and psychographic data to edit the requested Web page. The user profiles are preferably based on actual user Web use and surfing activity. Generating the user profiles preferably requires no or limited direct input from the users. The content of a Web page as published by a host Web server may be profiled to correlate components of the Web page with demographic and psychographic data or other data related to the user profiles. The user profiles may then be used to filter the content of the profiled Web pages for delivery to requesting users. A proxy server monitors user requests made through their Web clients, and filters the content of the requested page based on the user profile and the Web page profile, before delivering the page to the user. The system preferably provides to the requesting user, through a proxy server, an edited version of the HTML file for the original published Web page that is served by the host Web server. In embodiments of the invention, a single proxy server or proxy server system personalizes Web pages from multiple Web servers, using a single user profile for an individual user.

These and other features and advantages of the present invention will become readily apparent from the following detailed description, wherein embodiments of the invention are shown and described by way of illustration of the best mode of the invention. As will be realized, the invention is capable of other and different embodiments and its several details may be capable of modifications in various respects, all without departing from the invention. Accordingly, the drawings and description are to be regarded as illustrative in nature and not in a restrictive or limiting sense, with the scope of the application being indicated in the claims.

Brief Description of the Drawings

For a fuller understanding of the nature and objects of the present invention, reference should be made to the following detailed description taken in connection
5 with the accompanying drawings, wherein:

Figure 1 is a screen-shot in a browser window of a representative Web page, a home page provided by Yahoo!®, a typical Web content provider.

Figure 2 is a screen-shot in a browser window of a Yahoo!® Web page that allows users to access a Web-site specific, explicit personalization feature.

10 Figure 3 is a screen-shot in a browser window of a Yahoo!® Web page that allows users to select specific types of content for a personalized "My Front Page."

Figure 4 is a screen-shot in a browser window of a Yahoo!® Web page that allows users to modify the layout of "My Front Page."

15 Figure 5 is a screen-shot in a browser window of a Yahoo!® Web page that allows users to select a particular display scheme for Yahoo!® Web pages.

Figure 6 is a screen-shot in a browser window of a sample Yahoo!® "My Front Page" for a sports fan.

Figure 7 is a block diagram illustrating a representative network in which the inventive system is preferably implemented.

20 Figure 8 is a block diagram illustrating an alternative representative network in which the inventive system is preferably implemented.

Figure 9 is a block diagram illustrating the proxy server component of the inventive system.

25 Figure 10 is a screen-shot in a browser window of a representative Web page that has been personalized in accordance with one aspect of the inventive system.

Figure 11 is a screen-shot in a browser window of a representative Web page that has been personalized in accordance with a second aspect of the inventive system.

Detailed Description of the Preferred Embodiments

5 Figure 7 illustrates a representative network environment in which the inventive system may be implemented, with a first system architecture. Fig. 8 illustrates an alternative representative network environment in which the inventive system may also be implemented, with an alternative system architecture. Although
10 the inventive system is described herein primarily with reference to the system architecture of Fig. 7, the inventive system may also be implemented in accordance with Fig. 8.

Embodiments of the present invention are directed to providing personalization of Web content in real time to meet the interests of requesting users.

15 The network 100 may provide users with access to remote servers through the medium of the Web. The Web is a multimedia information retrieval system for accessing electronic information, typically via the Internet. In particular, the "Web" may refer to a collection of servers of the Internet that interact using the Hypertext Transfer Protocol (HTTP). The HTTP application protocol provides users access to
20 files on those servers that are defined using, e.g., a standard page description language known as Hypertext Markup Language (HTML). "Web pages" are files defined in the HTML format and can incorporate or link to different file formats such as text, graphics, software, audio, video, etc.

25 The network 100 includes a plurality of client machines 110 operated by various individual users to access the files over the network 100. A client machine 110 may be operated by one or more users. The client machines connect to multiple

servers 118 via communication channel 120, which is preferably the Internet.

Communications channel 120 may, however, alternatively comprise an intranet or other known networks or connections. In the case of the Internet, the servers 118 are Web servers that are supported by Web content providers and that are accessible by various clients.

The Web servers 118 operate or host so-called "Web sites" and support HTML files in the form of "Web pages" and documents (including text files, graphics files, software files, video files, audio files, etc.) in various formats linked to the Web pages. HTML provides basic document formatting for the Web pages and allows developers to specify links from the Web pages to other servers 118 and files. These links may be specified as "hyperlinks," which are text phrases or graphic objects that conceal the address of a site on the Web. The main page provided on a Web site typically provides access to various types or classes of information on that Web site, on other Web pages, or possibly on other Web sites, and is referred to as a "home page." A network path to a Web site or a Web page supported by a server 118 is identified by a Uniform Resource Locator (URL).

Users access Web pages of Web sites hosted on the Web servers 118 by specifying the URLs of the desired Web pages at the client computers 110. One example of a client machine 110 is a personal computer such as a Pentium-based desktop or notebook computer running a Windows operating system. A representative computer includes a computer processing unit, memory, a keyboard, a mouse and a display unit. The screen of the display unit is used to present a graphical user interface (GUI) for the user. The GUI is supported by the operating system and allows the user to use a point and click method of input, e.g., by moving the mouse pointer on the display screen to an icon representing a data object at a particular

location on the screen and pressing on the mouse buttons to perform a user command or selection. Also, one or more “windows” may be opened up on the screen independently or concurrently, as desired. A client machine 110 may also include, for example, a personal digital assistant, a handheld wireless telephonic device, or any other network access device.

Client machines 110 are enabled to access servers 118, interact over the Web and display Web content by known software tools called “browsers.” Representative browsers include, among others, Netscape® Navigator® and Microsoft® Internet Explorer®. A user of a client machine 110 having an HTML-compatible browser (such as Netscape® Navigator®) can retrieve a Web page (namely, an HTML formatted document) of a Web site by specifying the URL (e.g., www.yahoo.com) in an HTTP request that is sent over the Internet. Upon such specification, the client machine 110 makes a transmission control protocol/Internet protocol (TCP/IP) request to the server 118 identified in the link and receives the Web page in return.

Client machines 110 usually access servers 118 through some private Internet service provider (ISP) such as, e.g., America Online. Illustrated in Figure 7 is the ISP “point-of-presence” (POP), which includes an ISP POP server 112 linked to the client machines 110 for providing access to the Internet. The POP server 112 is connected to a section of the ISP POP local area network (LAN) that contains the user-to-Internet traffic. As described in the ‘755 application, the ISP POP server 112 may capture URL page requests from individual client machines 110 for use in user profiling and also distributes retrieved Web pages to users.

As discussed above, the inventive system is a method and system for dynamically personalizing published Web pages available on Web servers on the Internet for delivery to requesting users of the Web. The inventive system tailors the

content of published Web pages in accordance with a profile of the requesting user at the time the request is made and delivers a personalized HTML file to the user.

Preferably, the inventive system incorporates a proxy server component 114 linked to the ISP POP server 112 that handles the personalization function. In the inventive

5 system, a proxy server 114 fulfills user HTTP requests with Web pages personalized to the requesting users' profiles, when appropriate. Generally, the proxy server 114 monitors HTTP requests made by users, retrieves the requested Web pages, modifies the Web pages in accordance with a profile of the requesting user, and provides the

modified Web pages to the users through the POP server 112. Also, as will also be
10 discussed in detail below, the inventive system may further include a master server 116 linked to the proxy server 114 and the ISP POP server 112 through the Internet 120. The master server 116 handles administration and synchronization functions.

The system software is preferably distributed over the network 100 at the ISP POP server 112, the proxy server 114, and the master server 116 as will be discussed

15 below. The network environment may further include, for example, other components and system software for profiling (not shown herein) as discussed in the '755 application.

As shown in Fig. 7, the proxy server 114 is preferably directly linked between the POP server 112 and the Internet 120. In this case, the proxy server 114 functions
20 as a gateway for HTTP requests made by clients 110 of the POP server 112.

Alternatively, as shown in Fig. 8, a proxy server 114 may be indirectly linked to the POP server 112 by the Internet. In this case, HTTP requests are transmitted to the proxy server 114 from the POP server 112 via the Internet 120 using standard TCP/IP protocols. A single proxy server may handle HTTP requests from more than one POP
25 server; conversely, multiple proxy servers may handle HTTP requests from a single

POP server. Alternatively, the proxy server may be eliminated and its functionality incorporated in each POP server.

Fig. 9 further illustrates the proxy server 114 that accomplishes the personalization of delivered Web pages in preferred embodiments of the inventive system. The proxy server 114 may include a request generation component 122 and a Web page personalization component 124. In order to process an HTTP request for delivery of a personalized Web page, the request generation component 122 prepares valid requests for the Web page personalization component 124. For example, the request generation component 122 may initially obtain the request, check that the requested Web page is subject to personalization by the proxy server 114 and associate the request with a user profile. The request generation component 122 may also, for example, retrieve a Web page from a Web server or locate a profile for a Web page.

The request generation component 122 obtains HTTP requests, i.e., URLs, that are outgoing from the clients 110 to the Web servers 118 over the Internet 120. The request generation component 122 may obtain HTTP requests by monitoring all traffic outgoing from the POP server 112 to the Internet 120 with a sniffer to detect outgoing Web page requests. When the sniffer detects an outgoing Web page request from a client 110, it captures the associated packets and extracts the actual URL. User-to-Internet traffic that does not contain an HTTP request passes through the proxy server 114.

Web content providers may request that certain Web pages on their Web servers 118 be personalized or not be personalized in the inventive system, and may specify certain preferences or requirements or other processing instructions regarding the handling of the personalization. For example, Web content providers may prefer

that only Web site home pages be personalized, because subsequent link selections by users will inherently ensure that deeper content is of interest to the user. Web content providers may have proxy server accounts that maintain information regarding these issues, stored in account information database 126. In order to process an HTTP request, the request generation component 122 also determines whether the HTTP request is for a Web page subject to personalization, and the corresponding instructions, if any. If the requested Web page is not subject to personalization, the HTTP request may pass through the proxy server 114.

The sniffer of the request generation component also extracts information that may be used to correlate the URL request with a particular user profile. User profiles may typically be stored by reference to anonymous user IDs. So, for example, the sniffer may extract the client's IP address and cross-reference an anonymous user ID table provided by the POP server 112 to obtain the appropriate anonymous user ID for an HTTP request. If multiple users share a single client 110, then each user may be requested to register and to log in at the initiation of a Web session. That log in information may be associated with a user's HTTP request to be further associated with the anonymous user ID that is used to reference the user's profile. Under some circumstances, for example, if the request generation component 122 determines that requesting user does not have an anonymous user ID and/or user profile, the associated HTTP request may pass through the proxy server 114. The user ID information may be stored locally in user profile database 128 or remotely, e.g., at the master server 116 or at the POP server 112. User profile information may be synchronized by the master server 116 periodically, if stored locally.

Alternative to the sniffer in the request generation component 122 of the proxy server 114, the POP server 112 may direct user-to-Internet traffic containing HTTP

requests for processing by the proxy server 114 and direct other user-to-Internet traffic to bypass the proxy server 114. The POP server may also associate HTTP requests directed to the proxy server 114 with a user ID so that the proxy server 114 need not determine that information. The proxy server 114 may still confirm that the requested Web page is subject to personalization and obtain any corresponding processing instructions.

The personalization component 124 uses the user profile and a profile of the HTML file to edit the HTML file for the Web page. When the request generation component 122 generates an HTTP request that is eligible for personalization and associated with a user ID, the Web page personalization component 124 accesses the user profile and a profile of the HTML file for the requested Web page, analyzes the data to match the Web page content to user preferences, and produces a modified version of the HTML file for the Web page, personalized in accordance with the user profile. The proxy server 114 preferably obtains the profile from a local user profile database 128. Each user profile may contain, for example, demographic and psychographic data. For example, a user profile may take the following form:

User ID	Sports	Finance	Movies	Music	TV	...	Health	Gardenin g
1	10.0 (.75)	21.1 (.62)	0.0 (1.00)	9.4 (.84)	0.0 (1.00)	...	50.0 (.77)	85.0 (.82)

In the illustrated example, each psychographic category in the profile includes an affinity rating, on a scale of 0.0 to 100.0, followed by a confidence measure for that affinity rating. Each user profile is preferably generated by tracking the user's actual Web surfing activity and analyzing the user's click-stream data, as described in the '755 application.

A profile includes classifications for the content components of the HTML file for the requested Web page. The profile of the HTML file for the requested Web page is based on the same or a similar content classification scheme to the user profiles. An HTML file is formed of constituent components that include content components and formatting components, i.e., the HTML "mark-up." The content components include but are not limited to text, images, advertisements and links to other Web pages. By way of example, a content component can comprise the hyperlinked subject heading "Arts & Humanities" 21. The profile of the HTML file for the requested Web page preferably includes a content classification or affinity rating for each content component on the Web page that is subject to personalization. An HTML file profiler parses each HTML file to extract the constituent components, and analyzes and assigns ratings to the content components.

Content components may be associated with demographic and psychographic categories or assigned affinity ratings for a range of categories. Each content component may be evaluated, e.g., by matching keywords in text content components to content affinities or by translating URLs in Web page link content components to content affinities through a categorized URL database. Classification information, such as a categorized URL database, may be provided by entities such as Nielsen. Web content provider processing instructions may also be applied to or incorporated in the profile. Web content providers may also specify certain content affinities for content components of a page. Some content components on a Web page may not be subject to personalization, particularly if the Web content provider has specified that particular components should remain as is in the Web page delivered to the client; these components may be protected in the profile. Certain content components may

also be tied together such that if one is profiled, the other is profiled accordingly. Any other instructions from the Web content provider may also be tied into the profile.

After receiving a personalization request from the request generation component 122, the Web page personalization component 124 preferably obtains the associated user profile, HTML file and HTML file profile. The HTML file profiler may be located at the proxy server 114, or may be remotely located, for example, at master server 116. An HTML file may be obtained and profiled in advance, and the original file and the profile may be cached for access by the proxy server in an HTML file profile database 127, or may be dynamically profiled at the time an HTTP request for that HTML file is received from the client. Profiles may be generated by a combination of automated and manual profiling (e.g., by specific instructions supplied by the Web content provider). It is contemplated that an HTML file and its profile may be merged into one combined profiled version of the HTML file rather than maintained as two separate files. If the HTML file is not cached in advance, the proxy server 114 requests the Web page, obtains the HTML file and obtains the profile. If the HTML file is cached for use by the proxy server 114, the proxy server 114 preferably confirms that the cached file (and associated profile) is up-to-date and also transmits the HTTP request to the Web server 118 that originally served the page or maintains a record of the HTTP request so that the Web content provider can accurately register the number of hits to the page.

To personalize a requested Web page, the Web page personalization component 124 analyzes the respective user profile and HTML file profile to determine the most effective organization for the content of the requested Web page for display to that particular user. The proxy server 114 preferably accesses the profiled version of the HTML file from the HTML file profile database 127. In

accordance with the inventive system, the content of the Web page may be reorganized in several ways to produce a modified, personalized Web page. Certain content components, if deemed to be of low interest to the user, may be eliminated from the Web page display altogether. Generally, it is preferred to preserve access to all of the content of the original Web page. A link to "Other" content or a link to the original Web page may be provided and a message that the Web page has been personalized may be included in the modified HTML file to ensure that the user is able to access all of the content, if desired. Other content components may be rearranged to position content for which the user has a higher affinity so that it is more easily viewed, for example, by moving it to the top of a list, moving it "above the fold," or setting it apart so that it has more white space around it. Additional content may also be inserted if desired. For example, certain advertisements or links to articles may be included or excluded. Other advertisements or links to articles may be moved to better target the user's preferences. Content may also be modified so that the font or color or other graphics properties are changed.

The Web page personalization component 124 uses the classification of each content component from the profile to analyze its relevance to the requesting user. Content components may be matched to user profiles in any number of ways, for example, by using a certain threshold for the content affinity rating for a user to trigger content components corresponding to that content category. The proxy server 114 provides a modified Web page for display by creating a modified HTML file, with the included content components marked up with HTML code to specify the desired Web page display format.

Although a user profile database derived from clickstream data is preferably the main source for profile information, other sources of profile information may also

be employed. For example, geographic information may readily be inferred from a user's IP address, which is transmitted with every URL request. An ISP may also supply user ZIP codes, which provide an alternative means to geographically profile a user. Geographic data could also be included in a stored user profile as described
5 above. Geographic data may be used to deliver personalized content particular to a geographic area, such as local news and weather.

This modified HTML file is then forwarded to the client 110 through the POP server 112 for viewing by the user. When the revamped file is received at the client 110, the client browser interprets the HTML in the received HTML file and displays
10 the Web page for the user, just as it would have the original Web page from the original, published HTML file. A screen-shot of an exemplary personalized Web page 130 is shown in Fig. 10. The Yahoo!® home page 10 shown in Fig. 1 has been rearranged to better meet the interests of a hypothetical user. The header 12, quick access index 14, quick shopping index 16 and news sidebar 18 have not been
15 modified, for example, in accordance with Web content provider processing instructions with respect to certain constituent components of a Web page. However, the taxonomy-based directory 20 has been rearranged to put subject areas expected to be of greater interest to the user at the top of the list. For example, "Education" 23, "Reference" 25 and "Science" 26 have been moved up; and, "News & Media" 24,
20 "Arts & Humanities" 21, and "Business & Economy" 22 have been moved down.

A screen-shot of a second exemplary personalized Web page 140 based on the same Yahoo!® home page 10 is shown in Fig. 11. In personalized Web page 140, content not of interest to the user has been eliminated and the remaining content of interest to the user has been rearranged. Again, header 12, quick index 14, and news
25 sidebar 18 have not been edited. However, shopping quick index 16 has been edited

to eliminate Departments, Stores, and Features that are not of interest to the user based on his or her profile. In the taxonomy-based directory 20, categories of low interest to the user have been eliminated. Specifically, "Arts & Humanities" 21, and "Business & Economy" 22 have been eliminated, among others. This reorganization reduces what may be perceived by a user as clutter and greatly simplifies the presentation. However, the full functionality of the original page may be preserved, for example, by adding links to "Other" categories 142, as shown under Departments, and in the taxonomy-based directory. Thus, the user may more quickly and easily locate material that is most likely to be of interest to him or her and still access other areas of the Web site, when desired.

This rearrangement in accordance with the inventive system allows the user to more quickly and easily access the most pertinent subject areas for him or her. This rearrangement is transparent to the user, i.e., the user need not take specific steps to personalize or to invoke personalization of a particular Web page. Also, the user preferably receives the Web page without any perceptible delay as compared with regular delivery of a requested Web page. Moreover, if the content of a taxonomy-based directory changes, the user's preferred subject areas will continue to appear at the top of the taxonomy-based directory so long as his or her interests remain the same. Also, if the user's interests change, because the user profile is preferably tied to his or her Web surfing activity, those changes will automatically be recorded and taken into account without explicit action (e.g., changing selections on a checklist such as shown in Fig. 3) by the user.

The inventive system may also be combined with explicit preference selection by a user to enhance the automatic profiling. The present invention may be combined with selective delivery of advertising and other material as described in the '755

application. Although the inventive system has been described primarily with reference to an Internet-based network environment, the inventive system could also be implemented in a local-area network environment, for example. Also, while particular data structures, information storage and software distribution schemes have been described, any suitable scheme may be used. While the present invention has been illustrated and described with reference to preferred embodiments thereof, it will be apparent to those skilled in the art that modifications can be made and the invention can be practiced in other environments without departing from the spirit and scope of the invention, set forth in the accompanying claims.

We claim:

1. A method for personalizing a Web page with content for a user, comprising the steps of:

providing a profile of the Web page;

5 providing a profile of the user; and

producing a modified Web page based on the profile of the Web page and the profile of the user.

2. The method of claim 1, said profile of the user including demographic data.

3. The method of claim 2, the step of producing a modified Web page
10 including making a portion of the Web page content generally matching the demographic data more prominent.

4. The method of claim 2, the step of producing a modified Web page including making a portion of the Web page content not matching to the demographic data less prominent.

15 5. The method of claim 1, said profile of the user including psychographic data.

6. The method of claim 5, the psychographic data including a set of content affinities, the set of content affinities including a subset of higher content affinities, the step of producing a modified Web page including making a portion of the Web
20 page content corresponding to the subset of higher content affinities more prominent.

7. The method of claim 5, the psychographic data including a set of content affinities, the set of content affinities including a subset of lower content affinities, the step of producing a modified Web page including making a portion of the web page content corresponding to the subset of lower content affinities less prominent.

8. The method of claim 1, said profile of the user including geographic data.

9. The method of claim 8, said geographic data being inferred from an IP address or a ZIP code.

10. The method of claim 8, the step of producing a modified web page
5 including providing Web page content matching the geographic data.

11. The method of claim 10, wherein the matching Web page content is news or weather information.

12. The method of claim 1, wherein the step of producing a modified Web page includes rearranging a portion of the content of the Web page.

10 13. The method of claim 12, wherein the rearranged portion includes links.

14. The method of claim 12, wherein the rearranged portion includes advertisements.

15. The method of claim 12, wherein the rearranged portion includes images.

16. The method of claim 12, wherein the rearranged portion includes text.

15 17. The method of claim 1, wherein the step of producing a modified Web page includes eliminating a portion of the content of the Web page.

18. The method of claim 17, wherein the eliminated portion includes links.

19. The method of claim 17, wherein the eliminated portion includes advertisements.

20 20. The method of claim 17, wherein the eliminated portion includes images.

21. The method of claim 17, wherein the eliminated portion includes text.

22. The method of claim 17, wherein the step of producing a modified Web page further includes providing a link to the eliminated portion.

23. The method of claim 1, the Web page including a content component, said profile of the Web page associating said content component with demographic data,
5 said step of producing a modified Web page including the step of matching the profile of the user to the demographic data for the content component.

24. The method of claim 1, the Web page including a content component, said profile of the Web page associating said content component with psychographic data,
said step of producing a modified web page including the step of matching the profile
10 of the user to the psychographic data for the content component.

25. A method for profiling a Web page, comprising the steps of:

obtaining the HTML file for the Web page, the HTML file including one or more content components;

defining a classification scheme including one or more categories; and

15 parsing the HTML file for the Web page to identify the one or more content components; and

associating at least one of the one or more content components with at least one of the one or more categories.

26. The method of claim 25, wherein the one or more categories include
20 demographic categories.

27. The method of claim 25, wherein the one or more categories include psychographic categories.

28. The method of claim 25, further including the step of associating at least one of the one or more content components with at least one processing instruction.

29. The method of claim 25, wherein the classification scheme is related to a classification scheme for user profiles.

5 30. The method of claim 25, further including the step of generating a profile file for the Web page.

31. The method of claim 25, wherein the one or more content components include key words, the step of associating including using key words.

10 32. The method of claim 25, wherein said content components include links, the step of associating including using a URL database.

33. A method for personalizing for individual users in accordance with their requests a plurality of Web pages that are published on a plurality of Web content provider sites and accessible to a plurality of users, comprising the steps of:

providing profiles of the plurality of Web pages;

15 providing profiles of the plurality of users;

monitoring requests from the plurality of users;

detecting an individual request for a particular Web page;

obtaining the particular Web page;

obtaining a profile for the particular Web page;

20 obtaining a profile for the individual user;

producing a modified Web page based on the profile for the particular Web page and the profile for the individual user; and

sending the modified Web page for delivery to the individual user.

34. The method of claim 33, wherein the step of providing profiles of the plurality of Web pages includes dynamically generating the profile for the particular Web page after detecting an individual request for a particular Web page.

5 35. The method of claim 33, wherein the step of providing profiles of the plurality of Web pages includes generating and caching the profile for the particular Web page before detecting an individual request for the particular Web page.

36. The method of claim 33, wherein the step of providing profiles of the plurality of users includes tracking click-stream data of the plurality of users.

10 37. The method claim 33, wherein the step of producing a modified Web page includes matching the profile of the particular Web page to the profile of the user and rearranging one or more portions of the particular Web page accordingly.

38. The method of claim 37, wherein the step of matching the profile of the particular Web page to the profile of the user includes applying a threshold value to a
15 content affinity rating in the profile of the user.

39. The method of claim 37, wherein the step of rearranging one or more portions of the particular Web page includes eliminating one or more portions of the particular Web page.

40. The method of claim 33, further including the step of obtaining processing
20 instructions from the plurality of Web site content providers, including processing instructions for the particular Web page, wherein the step of producing a modified Web page further includes applying the processing instructions for the particular Web page.

41. The method of claim 33, wherein the profile of a Web page comprises profiles of individual content components, and wherein producing a modified Web page comprises rearranging the content components of the Web page to only or to more prominently display components having profiles matching the profile of the individual user.

42. The method of claim 41, wherein a content component comprises a hyperlinked subject heading.

43. The method of claim 41, wherein a content component comprises an advertisement.

44. The method of claim 41, wherein a content component comprises an article.

45. A computer for personalizing Web pages in response to detecting user requests for the Web pages, comprising:

a memory for storing a program;

a processor operative with the program to:

(a) detect a request for a particular Web page by an individual user;

(b) obtain a profile of the individual user and a profile of the particular Web page;

(c) produce a modified Web page based on the profile of the individual user and the profile of the particular Web page; and

(d) send the modified Web page for delivery to the individual user.

46. The computer of claim 45, wherein said computer is a proxy server.

47. The computer of claim 46, wherein the program includes a profiler for producing a profile of the particular Web page.

48. A system for personalizing a Web page from a Web site of a Web content provider in response to a request by a user, comprising:

- 5 means for detecting a request for a Web page by a user;
- means for obtaining the Web page;
- means for obtaining a profile of the Web page;
- means for obtaining a profile of the user;
- means for producing a modified Web page based on the profile of the Web
- 10 page and the profile of the user; and
- means for delivering the modified Web page to the user.

49. The system of claim 48, further including means for generating a profile of the Web page.

50. The system of claim 49, further including means for caching a profile of

15 the Web page.

51. A system for personalizing a Web page from a Web site of a Web content provider in response to a request by a user, comprising:

- a first database containing profiles of a plurality of users;
- a second database containing profiles of a plurality of Web pages; and
- 20 a proxy server including a request generation component for processing a received request for a Web page and generating a valid personalization request and a personalization component for personalizing a Web page in accordance with a profile

of the user by generating a modified source file, the proxy server being linked to said first and second databases.

52. The system of claim 51, further comprising a profiler to generate the profiles of the plurality of Web pages.

5 53. The system of claim 51, wherein the proxy server is linked to a user computer for providing Web access to a user, the proxy server being linked to receive Web requests of the user and fulfill Web requests of the user.

54. The system of claim 53, wherein the proxy server is capable of handling Web requests to a plurality of Web sites.

1/11

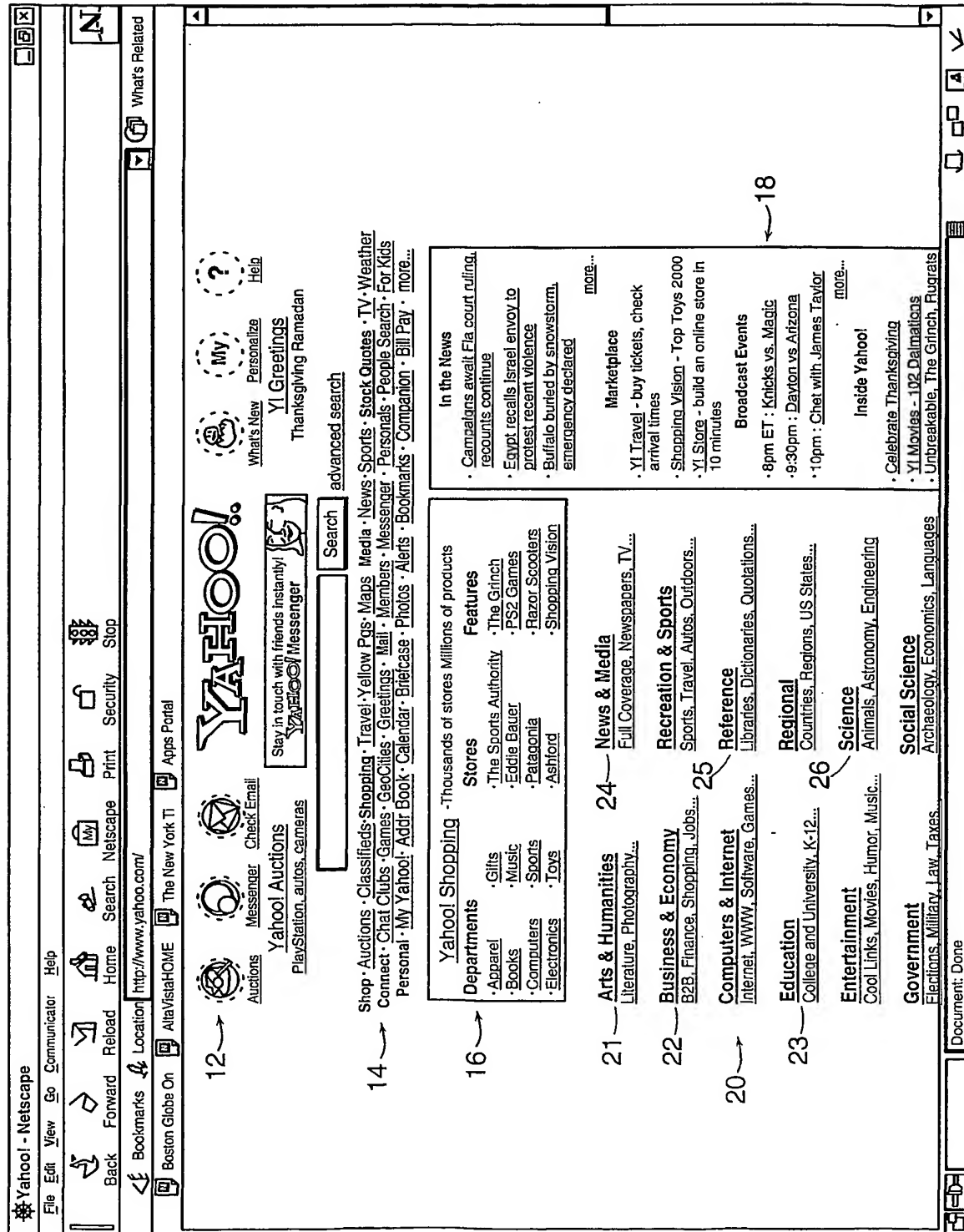
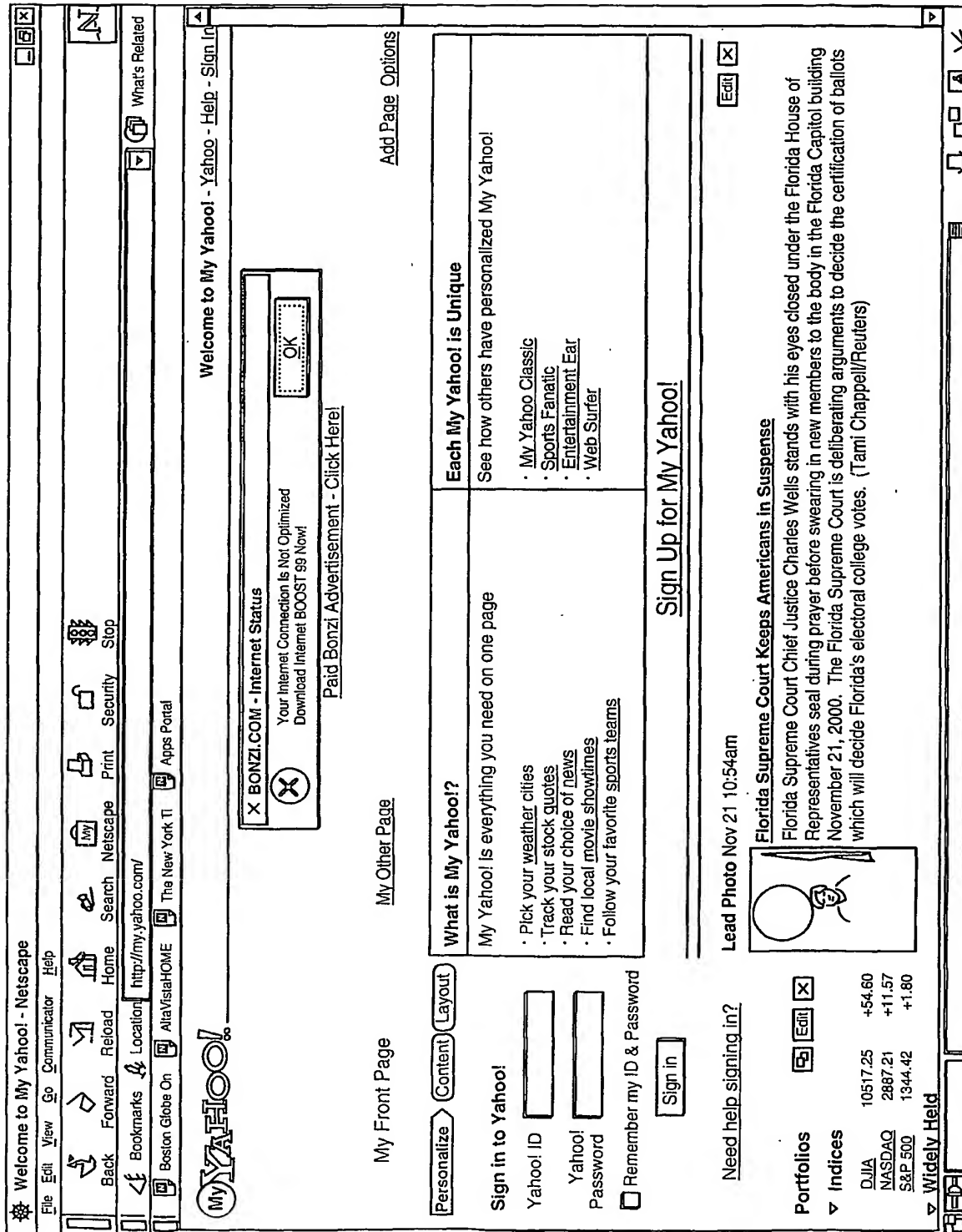


FIG. 1

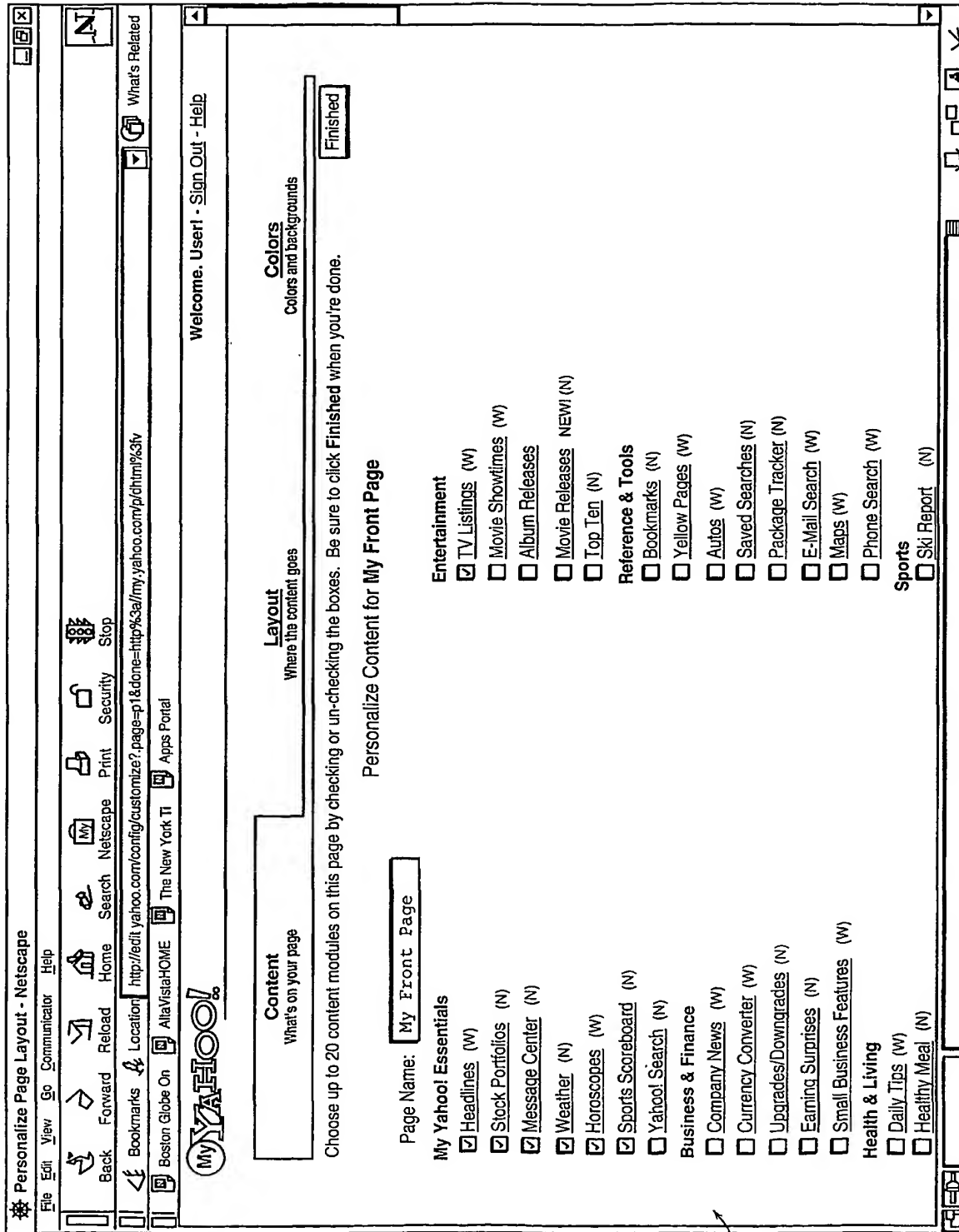
2/11

FIG. 2



3/11

FIG. 3



34

32

4/11

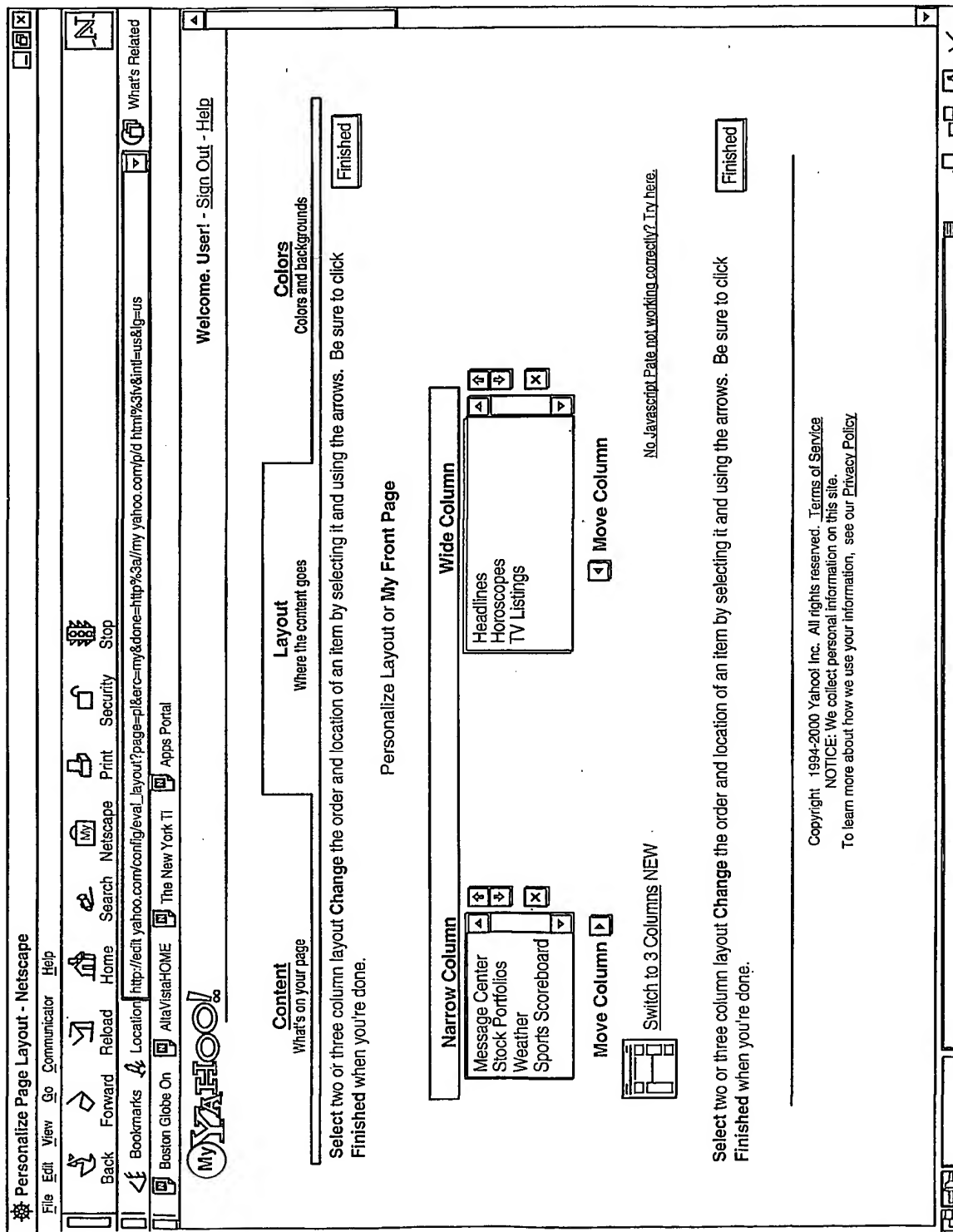


FIG. 4

5/11

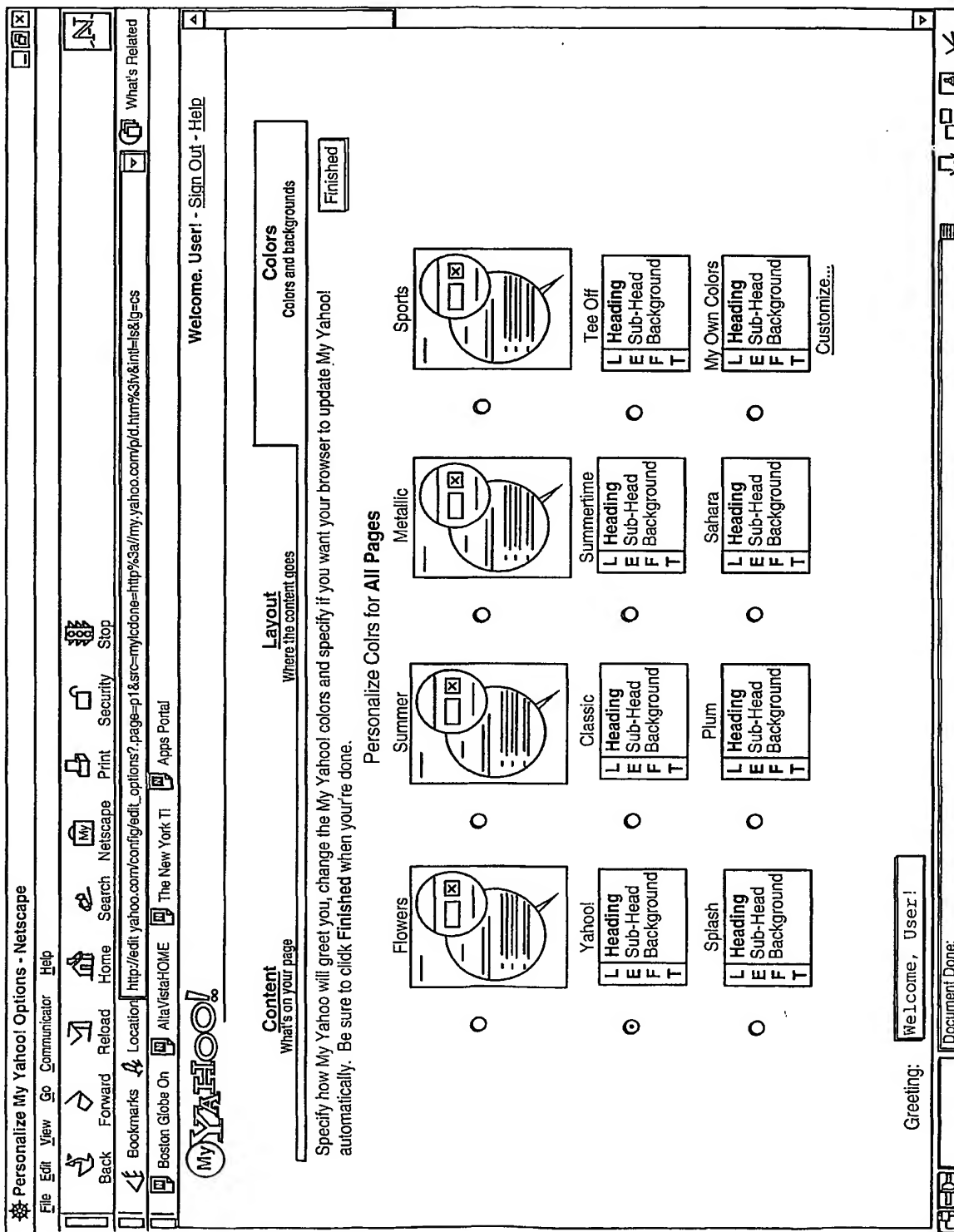


FIG. 5

6/11

Sport Guest - Netscape
File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location <http://my.yahoo.com/?pre=4>

Boston Globe On AlliaVisaHOME The New York Ti Apps Portal What's Related

Need help signing in? [My Sports Center Headlines Nov 22 1:43pm PT](#) [Edit](#) [X](#)

Customize your My Yahoo Headlines to see only the news of interest to you.

Scoreboard [Edit](#) [X](#) **Top Sports Stories from AP** [Edit](#) [X](#) Nov 22 1:35pm PT

TODAY
no games for selected teams

YESTERDAY
no games for selected teams

Weather [Edit](#) [X](#)

Los Angeles, CA 52..64 F

Miami, FL 50..72 F

New York, NY 30..36 F

San Francisco, CA 45..55 F

click on city for extended forecast
or search by Zip Code or City

[Search](#)

Major League Baseball [Edit](#) [X](#) Nov 22 1:26pm PT

- White Sox sign SS Valentin to three-year contract
- Padres claim Middlebrook off waivers
- Braves president suggests the team can't afford both A-Rod and Hampton

World Soccer [Edit](#) [X](#) Nov 22 1:09pm PT

- Leeds United 0 Real Madrid 2 -latest
- Bayern Munich 1 Olympique Lyon 0 - latest
- Brazil investigators seek source of fake passports

PGA [Edit](#) [X](#) Nov 22 11:08am PT

- Former Ryder Cup player misses card
- Baddeley has a pro's calm for Australian Open
- Singh looks to return to Masters victory form

Team News [Edit](#) [X](#)

Chicago Cubs [Edit](#) [X](#) Nov 22 12:12pm PT

- Sosa says talks ongoing with Cubs
- Cubs trade with Oakland for Matt Stairs
- Cubs acquire OF Stairs from A's for minor leaguer

New York Yankees [Edit](#) [X](#) Nov 22 12:12pm PT

Healthy Meal [Edit](#) [X](#)

Couscous with Sweet Corn
See Full Recipe

Health Tips [Edit](#) [X](#)

Fitness Tip

Document Done:

FIG. 6

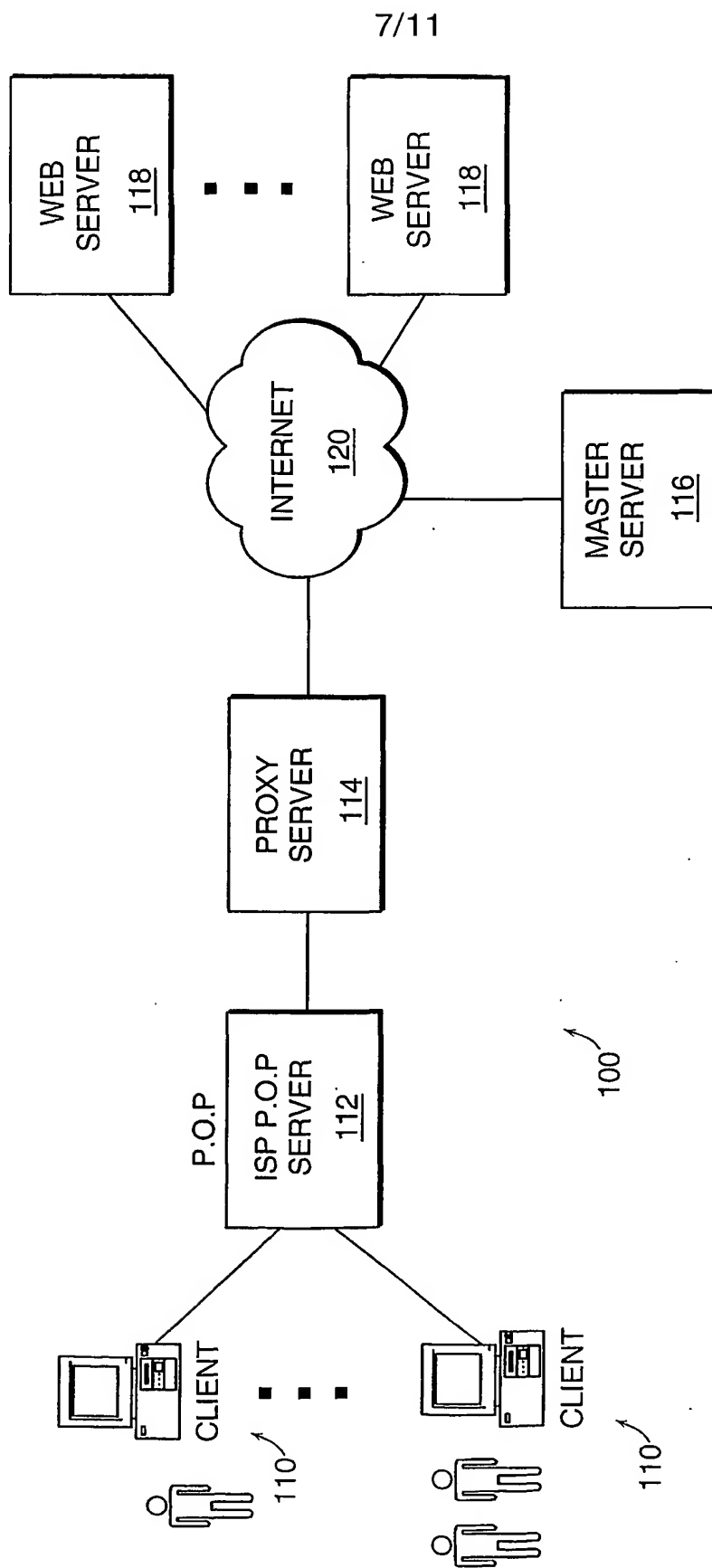


FIG. 7

8/11

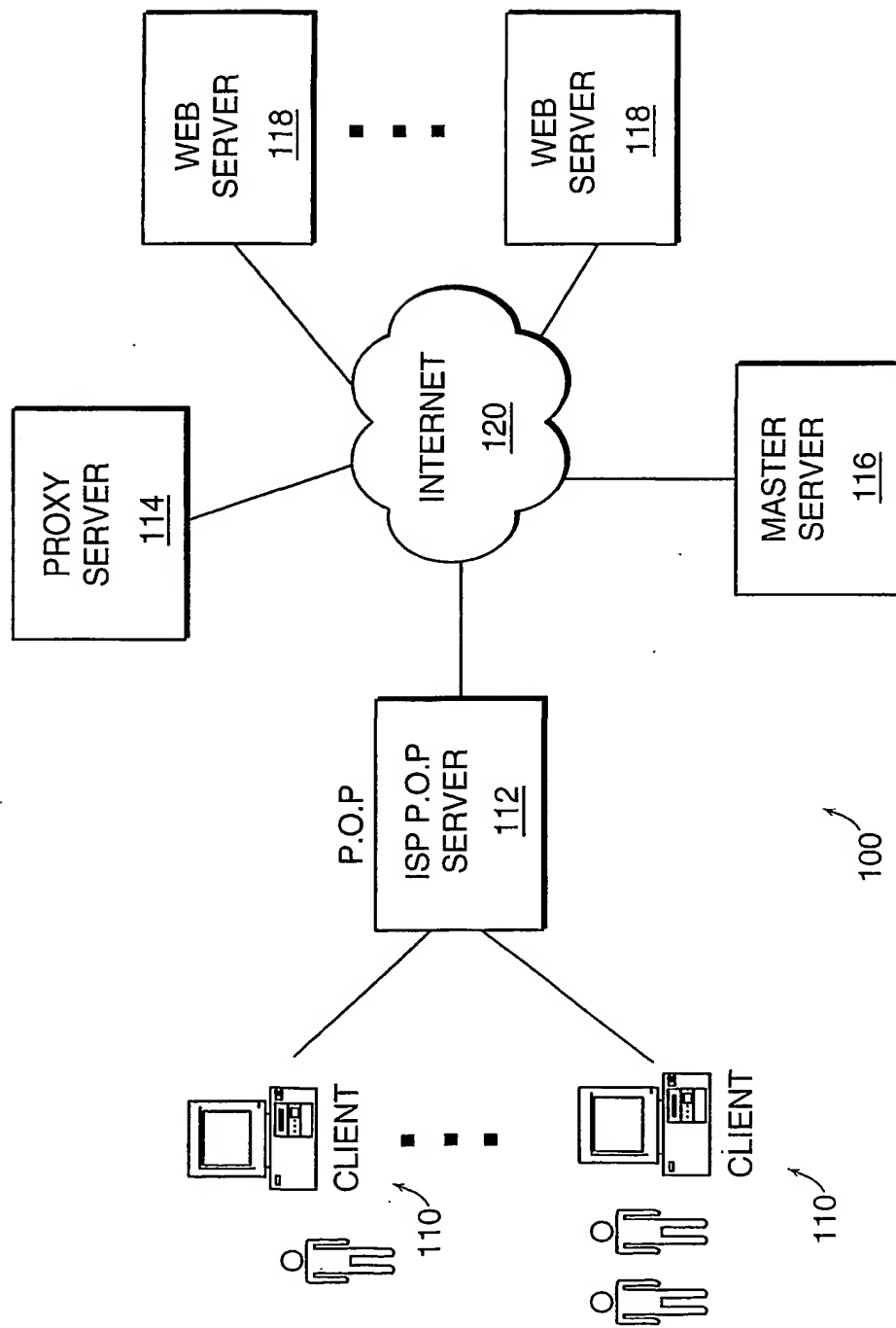


FIG. 8

9/11

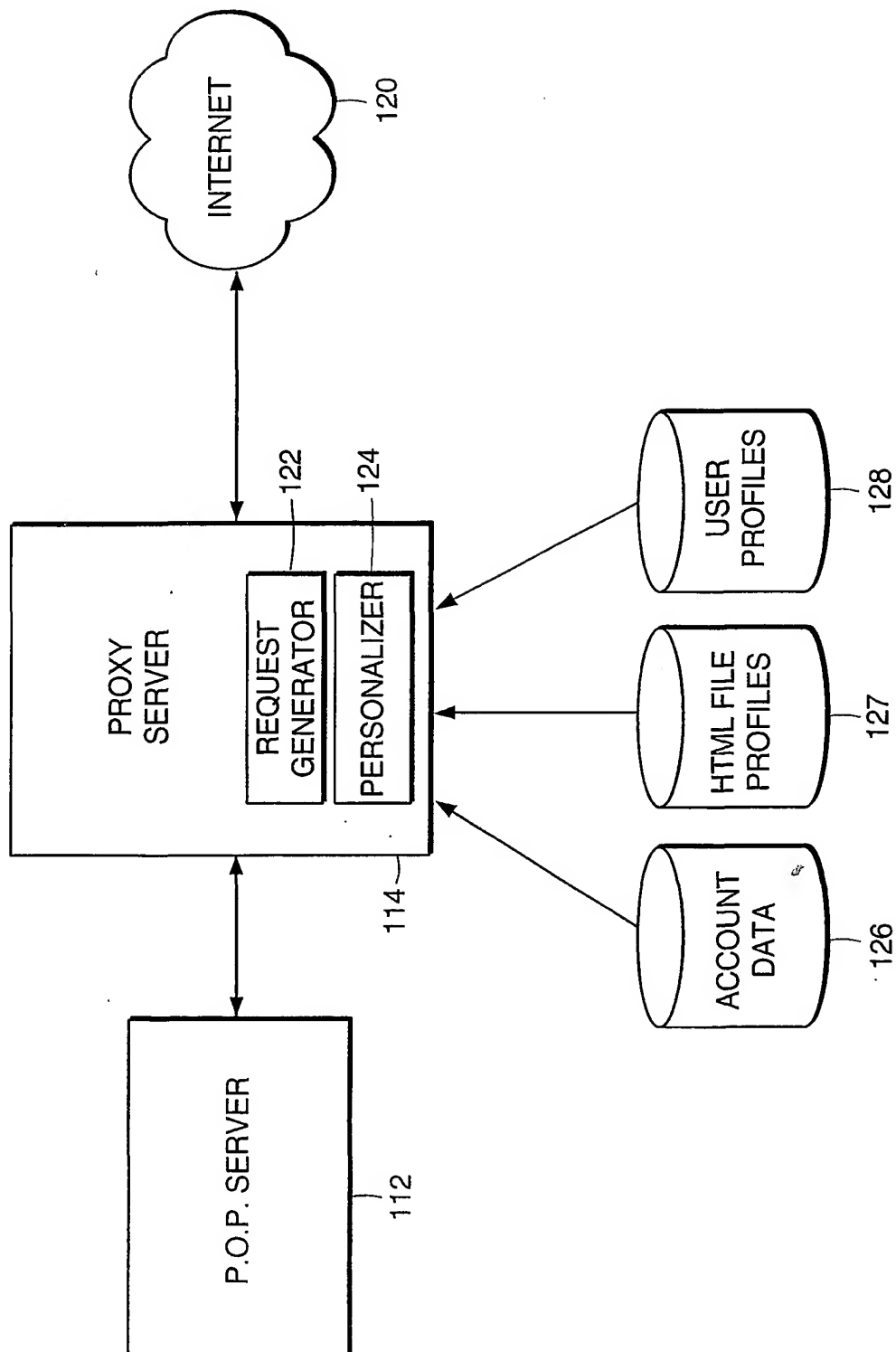
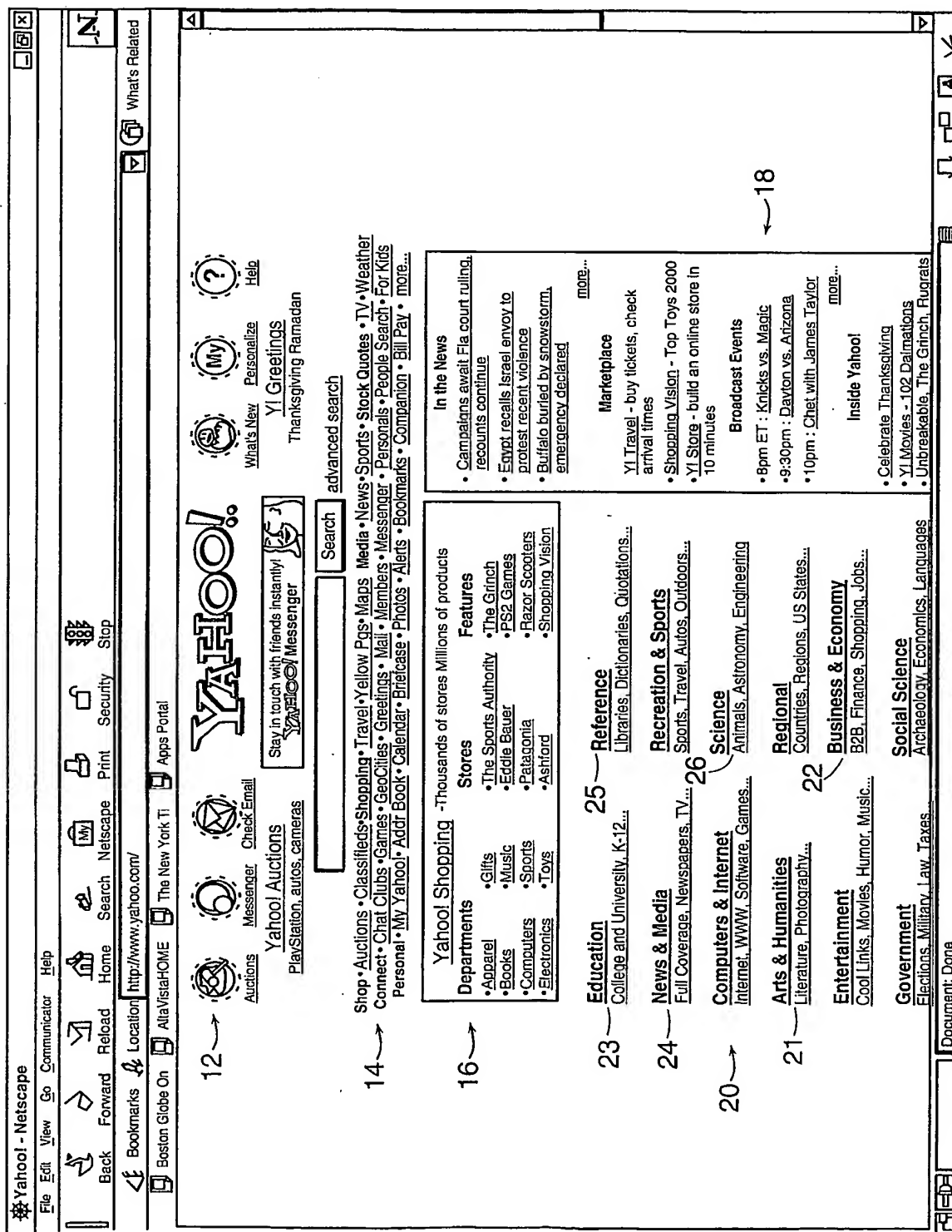


FIG. 9

10/11

FIG. 10



130-

11/11

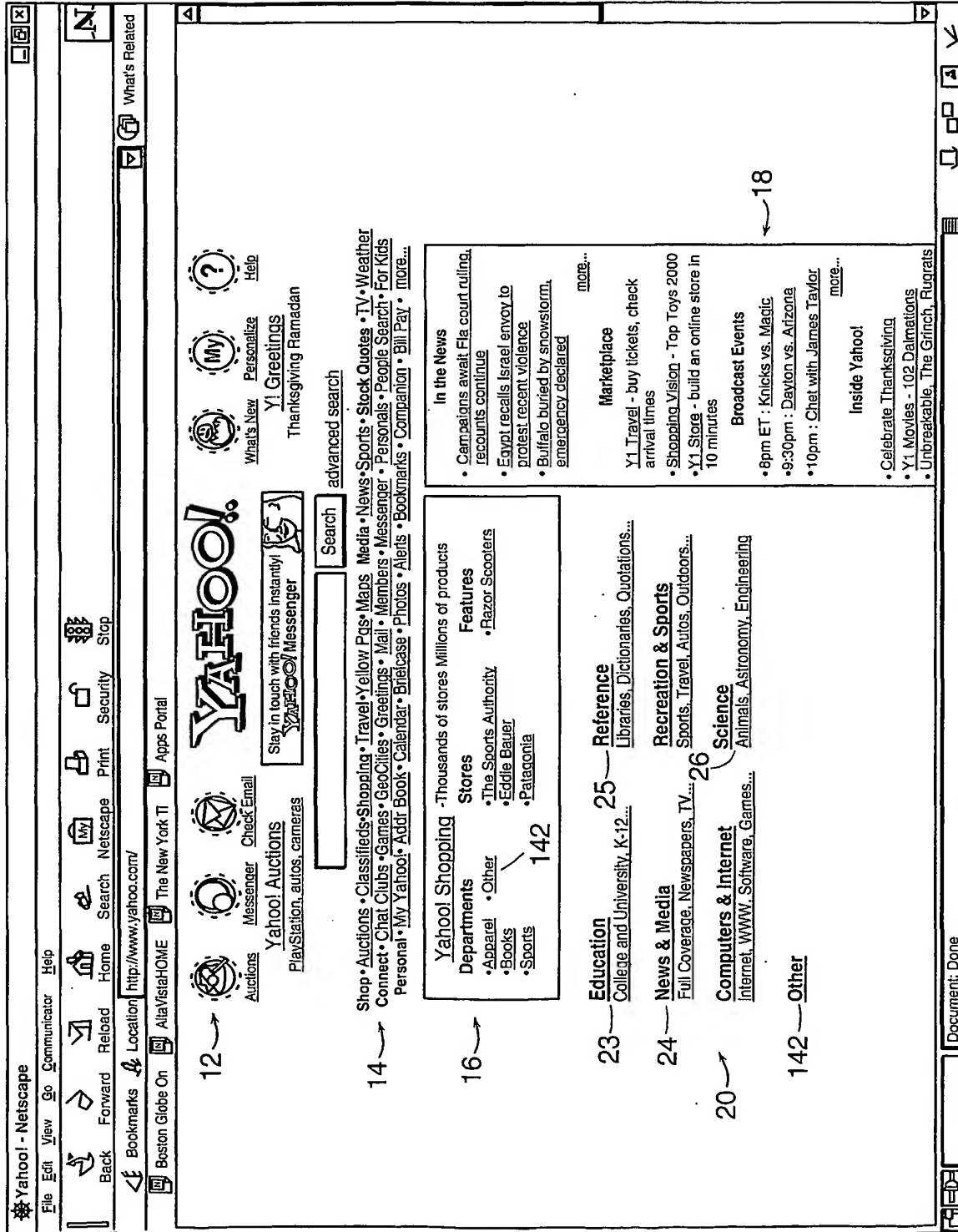


FIG. 11